


```

AC P09801;
DT 01-MAR-1989 (rel. 10, Created)
DT 01-MAR-1989 (rel. 10, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).
OS Gossypium hirsutum (upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RT Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XVII. cDNA and amino acid sequences of the members of
RT the storage protein families."
RL Plant Mol. Biol. 7:475-489(1986).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M16891; AAA3071.1; -.
CC DR PIR: A30838; FWCNAB.
CC DR HSSP: B50477; ICAX.
CC DR INTERPRO: IPR001113; -.
CC DR PFAM: PF00546; seedstore-7s; 1.
CC KM Seed storage protein; Signal.
CC FT SIGNAL 1 25
CC FT CHAIN 26 588
CC SQ SEQUENCE 588 AA; 69729 MW; 63BE99B29AB8ADEB CRC64;

Query Match 38.6%; Score 132; DB 1; Length 588;
Best Local Similarity 35.8%; Pred. No. 6.7e-06;
Matches 29; Conservative 11; Mismatches 19; Indels 22; Gaps 3;

OY 2 ORGYOCCGRCGCGOOGOREOQCCORCKWEQYKEQ-----ERGEHENVNH 46
DB 121 QRFQCGCHQCHQORRPRKQCCVACRECHRYQENPMWREREREAEETEBSGQDSHN 180
OY 47 ----HKK---NRSEEEGQOR 60
DB 181 PFHFHRSFQSRFHEHGNFR 201

RESULT 3
GBF_DICDI STANDARD: PRT; 708 AA.
ID GBF_DICDI
AC P36417;
DT 01-JUN-1994 (rel. 29, Created)
DT 01-JUN-1994 (rel. 29, Last sequence update)
DT 01-NOV-1997 (rel. 35, Last annotation update)
DE G-BOX BINDING FACTOR (GBF).
OS GBFA.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Dictyostelida; Dictyostelium.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RN STRAIN=AX3;
RX MEDLINE=94170994; PubMed=8125261;
RA Schmitzler G.R., Fischer W.H., Firtel R.A.;
RT "Cloning and characterization of the G-box binding factor, an
RT essential component of the developmental switch between early and
RT late development in Dictyostelium."

```

```

RL Genes Dev. 8:502-514(1994).
CC -1- FUNCTION: CAMP-RESPONSIVE TRANSCRIPTIONAL ACTIVATOR REGULATING
CC LATE GENE EXPRESSION. ESSENTIAL COMPONENT OF THE DEVELOPMENTAL
CC SWITCH BETWEEN EARLY AND LATE DEVELOPMENT. BINDS TO A NUMBER OF
CC CA/GT-RICH GENE REGULATORY ELEMENTS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L29075; AAA21021.1; -.
CC DR TRANSFAC: T00315; -.
CC DR DICTYDB: DD02046; GBFA.
CC KW Transcription regulation; Activator; DNA-binding; Nuclear protein;
CC Repeat.
CC FT REPEAT 339 368 1.
CC FT REPEAT 481 510 2.
CC FT DOMAIN 11 21 POLY-SER.
CC FT DOMAIN 115 263 GIN-RICH.
CC FT DOMAIN 270 292 POLY-ASN.
CC FT DOMAIN 549 557 POLY-ASN.
CC SQ SEQUENCE 708 AA; 79268 MW; B4BD8F04FACACCA CRC64;

Query Match 27.2%; Score 93; DB 1; Length 708;
Best Local Similarity 31.0%; Pred. No. 0.04;
Matches 18; Conservative 14; Mismatches 26; Indels 0; Gaps 0;

OY 2 ORGYOCCGRCGCGOOGOREOQCCORCKWEQYKEQERGEHENVNHKRNSEEEQ 59
DB 190 QH0HQ0Q0Q0Q0H0Q0H0Q0Q0Q0H0Q0H0Q0H0Q0H0Q0H0Q0H0Q0H0Q0Q 247

RESULT 4
INVO_CANFA STANDARD: PRT; 285 AA.
ID INVO_CANFA
AC P18174;
DT 01-NOV-1990 (rel. 16, Created)
DT 01-NOV-1990 (rel. 16, Last sequence update)
DT 01-FEB-1996 (rel. 33, Last annotation update)
DE INVOUCURIN.
GN IVL.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90348475; PubMed=2385171;
RA Tseng H., Green H.;
RT "The involucrin genes of pig and dog: comparison of their segments of
RT repeats with those of prosimians and higher primates."
RL Mol. Biol. Evol. 7:293-302(1990).
CC -1- FUNCTION: INVOUCURIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS
CC IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
CC MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE
CC FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC -1- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND
CC OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M34442; AAA30853.1; -.

```

```

OY 2 ORGYOCCGRCGCGOOGOREOQCCORCKWEQYKEQERGEHENVNHKRNSEEEQ 59
DB 190 QH0HQ0Q0Q0Q0H0Q0H0Q0Q0Q0H0Q0H0Q0H0Q0H0Q0H0Q0H0Q0H0Q0Q 247

RESULT 4
INVO_CANFA STANDARD: PRT; 285 AA.
ID INVO_CANFA
AC P18174;
DT 01-NOV-1990 (rel. 16, Created)
DT 01-NOV-1990 (rel. 16, Last sequence update)
DT 01-FEB-1996 (rel. 33, Last annotation update)
DE INVOUCURIN.
GN IVL.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90348475; PubMed=2385171;
RA Tseng H., Green H.;
RT "The involucrin genes of pig and dog: comparison of their segments of
RT repeats with those of prosimians and higher primates."
RL Mol. Biol. Evol. 7:293-302(1990).
CC -1- FUNCTION: INVOUCURIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS
CC IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
CC MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE
CC FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC -1- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND
CC OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M34442; AAA30853.1; -.

```

[illegible]

```
CC or send an email to license@sb.stl.ch).
DR EMBL; X67204; CAB56798.1; -
CC DR EMBL; U03645; AAB60446.1; -
DR DR EMBL; X55491; CAA39111.1; -
CC HSSP; O05066; IHRZ.
DR TRNSFPAC; T00996; -
DR MCD; MGI; 98660; TDY.
DR INTERPRO; IPRO00910; -
CC PFAM; PF00505; HMG_box; 1.
KW DNA-binding; Nuclear protein; Transcription regulation; Activator;
KW Sexual differentiation; Repeat.
FT DNA_BIND 5 73 HMG_BOX.
FT VARIANT 63 63 W -> T (IN STRAIN TORINO).
FT VARIANT 133 133 I -> L (IN STRAIN TORINO).
FT VARIANT 143 145 LOO -> P (IN STRAIN TORINO).
FT VARIANT 169 170 MISSING (IN STRAIN TORINO).
FT VARIANT 209 209 H -> Q (IN STRAIN TORINO).
FT VARIANT 211 211 E -> Q (IN STRAIN TORINO).
FT VARIANT 235 395 MISSING (IN STRAIN TORINO).
SQ SEQUENCE 395 AA; 49495 MM; FEEB9C3516ICDB0C CRC64;

Query Match 26.5%; Score 90.5; DB 1; length 395;
Best Local Similarity 28.6%; Pred. No. 0.041; 5; Gaps
Matches 18; Conservative 20; Mismatches 20; Indels 5;

QY 3 ROYOSCGRCEDDGGREDDOC-----ORKSCDEKCEDEGENENHNHKKRSEEEEG 57
Db 160 000000GFNDNNQQ0000000000GFNDNNDQKQGFNDNNDQ0000000000GFNDNN 219
QY 58 QQR 60
Db 220 QQQ 222

RESULT 5
SNF5_YEAST STANDARD; PRT; 905 AA.
ID SNF5_YEAST AC P18480;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRANSCRIPTION REGULATORY PROTEIN SNF5 (SWI/SNF COMPLEX COMPONENT SNF5 DE (TRANSCRIPTION FACTOR TYE4)).
GN SNF5 OR TYE4 OR SWI10 OR YBR289W OR YBR2036.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-MCY;
RX MEDLINE=91042489; PubMed=2233708;
LA Laurent B.C., Treitel M.A., Carlson M.;
RT "The SNF5 protein of Saccharomyces cerevisiae is a glutamine- and proline-rich transcriptional activator that affects expression of a broad spectrum of genes.";
RL Mol. Cell. Biol. 10:5616-5625(1990).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-S288C;
RX MEDLINE=94378722; PubMed=8091861;
LA Holmstrom K., Brandt T., Kallioes T.;
RT "The sequence of a 32,420 bp segment located on the right arm of chromosome II from saccharomyces cerevisiae.";
RL Yeast 10:S47-S62(1994).
CC -|- FUNCTION: INVOLVED IN TRANSCRIPTIONAL ACTIVATION. THE SWI/SNF COMPLEX IS REQUIRED FOR THE INDUCED EXPRESSION OF A LARGE NUMBER OF GENES. THIS COMPLEX ALTERS CHROMATIN STRUCTURE TO FACILITATE BINDING OF GENE-SPECIFIC DEDICATED TRANSCRIPTION FACTORS.
CC -|- SUBUNIT: COMPONENT OF THE SWI/SNF GLOBAL TRANSCRIPTION ACTIVATOR COMPLEX.
CC -|- SUBCELLULAR LOCATION: NUCLEAR.
```

CC	-	SIMILARITY: BELONGS TO THE SNR5 FAMILY.
CC	----	
CC	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	CC	the European Bioinformatics Institute. There are no restrictions on its use by
CC	CC	non-profit institutions as long as its content is in no way
CC	CC	modified and this statement is not removed. Usage of it for commercial
CC	CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).
CC	CC	----
DR	EMBL; M36482; AAA35062.1; -	
DR	EMBL; X76053; CAA33652.1; -	
DR	EMBL; Z36158; CAA85254.1; -	
DR	PIR; S44551; RGRYS.	
DR	PIR; S39145; S39145.	
DR	SGD; S0000493; SNF5.	
KW	Transcription regulation; Activator; Nuclear protein.	
FT	DOMAIN	31 270
FT	DOMAIN	72 132
FT	DOMAIN	272 324
FT	DOMAIN	489 588
FT	DOMAIN	714 882
FT	DOMAIN	755 798
FT	CONFLICT	564 564
SO	SEQUENCE	905 AA: 102557.MW: A287B4A648DD1A5 CRC64;

	Query Match	26.3%	Score 90	DB 1	Length 905
	Best Local Similarity	29.1%	Pred. NO.	0.096	
	Matches 16,	Conservative	20,	Mismatches 19,	Indels 0,
	Gaps	0,			
OY	4 OYOCGCGCQEDQDQCQRCDQCCKRCWCMDYKCEDEGEHENTHHNKKNSSEEGQ	58			
Db	216 QIQQQQQQKQKQQQQQQQHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ	270			

RESULT	7
ID	U2R2_HUMAN
NAME	U2R2_HUMAN
STANDARD	PRT; 482 AA.
AC	Q15696;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	15-DEC-1998 (Rel. 37, Last annotation update)
DE	U2 SMALL NUCLEAR RIBONUCLEOPROTEIN AUXILIARY FACTOR 35 KDA SUBUNIT
DE	RELATED-PROTEIN 2.
GN	U2AF1-RS2 OR U2AF1RS2.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=BRAIN;
RX	MEDLINE=96163878; PubMed=8586425;
RA	Kitagawa K., Wang X., Hatada I., Yamaoka T., Nojima H.,
RA	Inazawa J., Abe T., Mitsuya K., Oshimura M., Murata A., Monden M.,
RA	Mukai T.;
RT	"Isolation and mapping of human homologues of an imprinted mouse gene
RT	U2af1-rs1.";
RL	Genomics 30:257-263(1995).
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC	-1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC	-1- SIMILARITY: TO MAMMALIAN SPLICING FACTOR U2AF 35 KDA SUBUNIT.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; D49677.1; BAA08533.1; -.
DR	MI; 300028; -.

DR	INTERPRO: IPR000504; -	
DR	INTERPRO: IPR000571; -	
DR	PFAM: PF000476; rrm: 1.	
DR	PFAM: PF00642; zf-CCH: 1.	
DR	PROSITE: PS50102; RRM: 1.	
KM	Nuclear protein: Ribonucleoprotein.	
FT	DOMAIN	46
FT	DOMAIN	118
FT	DOMAIN	123
QO	SEQUENCE	482 AA; 58044 MW; 1DACCBACCA4727266 CRC64;

```

Query Match      25.4%  Score 87;  DB 1;  Length 482;
Best Local Similarity 25.4%  Pred. No. 0.11;
Matches 15;  Conservative 22;  Mismatches 22;  Indels 0;  Gaps 0;

OY      2  QRTATCCGCGQEDQGGQREDDQCCRCMKMEDYKEDERGENHYNHNKKNRSEEDGQOR 60
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      80  QKADEEFPIKKEEAAKKRRQEDERKRTKLEEMWEDQKKREFEEDQKKRQKKKEKALQK 138

```

RESULT	8	STANDARD:	PRT:	1154 AA.
CC	WCI_NEUCR			
AC	WCI_NEUCR			
DT	Q01371:			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	WHITE COLLAR 1 PROTEIN (WCI1).			
GN	WCI1.			
OS	Neurospora crassa.			
OC	Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=74-OR23-1A;			
RX	MEDLINE=96203083; PubMed=8612589;			
RA	Ballario P., Vittorioso P., Magrelli A., Talora C., Cabibbo A.,			
RT	Malino G.;			
RT	"White collar-1, a central regulator of blue light responses in			
RL	Neurospora, is a zinc finger protein."			
CC	EMBO J. 15:1650-1657(1996).			
CC	-1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR INVOLVED IN LIGHT			
CC	REGULATION. BINDS AND AFFECTS BLUE LIGHT REGULATION OF THE AL-3			
CC	GENE. WCI1 AND WC2 PROTEINS INTERACT VIA HOMOLOGOUS PAS DOMAINS,			
CC	BIND TO PROMOTERS OF LIGHT REGULATED GENES SUCH AS PRO, AND			
CC	ACTIVATE TRANSCRIPTION.			
CC	-1- SUBUNIT: HETERODIMER OF WCI1 AND WC2 (POTENTIAL).			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR.			
CC	-1- INDUCTION: BY BLUE LIGHT.			
CC	-1- DOMAIN: THE GLUTAMINE-RICH DOMAIN MIGHT FUNCTION IN ACTIVATING			
CC	GENE EXPRESSION.			
CC	-1- SIMILARITY: BELONGS TO THE GATA-TYPE TRANSCRIPTION FACTOR FAMILY.			
CC	-1- SIMILARITY: CONTAINS A PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X94300; CAA63964.1; -.			
DR	HSSP; P17678; 1GAU.			
DR	INTERPRO; IPR000014; -.			
DR	INTERPRO; IPR0000679; -.			
DR	INTERPRO; IPR001610; -.			
DR	PFAM; PF00320; GATA; 1.			
DR	PFAM; PF00785; PAC; 1.			
DR	PFAM; PF00989; PAS; 3.			
DR	PROSITE; PS00344; GATA_ZN_FINGER_1; 1.			
DR	PROSITE; PS00114; GATA_ZN_FINGER_2; 1.			
KW	Transcription regulation; Activator; DNA-binding; Zinc-finger;			
KW	Nuclear protein; Repeat.			

```

FT DOMAIN 16 61 GLN-RICH.
FT REPEAT 381 448 PAC-SL-1.
FT DOMAIN 469 508 PAC MOTIF.
FT REPEAT 576 642 PAS-2.
FT DOMAIN 650 691 PAC MOTIF.
FT REPEAT 695 761 PAS-3.
FT ZN_FING 935 960 GATA-TYPE.
FT DOMAIN 21 57 POLY-GLN.
FT DOMAIN 329 333 POLI-PRO.
SO SEQUENCE 1154 AA; 125944 MW; CBACBVB6047ECCDC5 CRC64;

Query Match      25.4%; Score 87; DB 1; Length 1154;
Best Local Similarity 37.3%; Pred. No. 0.23;
Matches 19; Conservative 13; Mismatches 15; Indels 4; Gaps 1;

Qy    1 LDRGYOOCGRCOEOREREOCQRSCRMEQRYKRGEGENHNHKKNR 51
     ||| | : : :||| :||| | :||| | :||| | :||| | :||| 
Db   15 LGNHGHNGHQDQQDDQQDDQQDQQD-----QDQDQDQHNDQQQTNQ 61


RESULT          9
LP61_EIMTE             STANDARD:           PRT:       255 AA.
ID LP61_EIMTE            AC PL5714;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE ANTIGEN IPMC-61 (FRAGMENT).
OS Eimeria tenella.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae; Eimeria.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=SPOROZOITE; MEDLINE=90348718; PubMed=2200963;
RA Ko C., Smith C.K., IT, McDowell W.;
RT "Identification and characterization of a target antigen of a monoclonal antibody directed against Eimeria tenella merozoites.";
RL Mol. Biochem. Parasitol. 41:53-64(1990).
CC -! FUNCTION: UNKNOWN. THE GLN-RICH TANDEM REPEATS MAY BE IMPORTANT FOR AN UNKNOWN ASPECT OF THE PARASITIC LIFE CYCLE. MAY BE AN IMPORTANT IMMUNOGEN.
CC -! SUBUNIT: MAY BE COVALENTLY LINKED BY DISULFIDE BONDS TO OTHER POLYPEPTIDES TO FORM THE 80 KDA ANTIGEN.
CC -! DEVELOPMENTAL STAGE: PRESENT IN ALL STAGES THROUGHOUT THE SPOROPLATION OF THE OOCYSTS AND IN THE SPOROZOOTES FOLLOWING EXCYSTATION.
CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).
CC DR EMBL; M30933; AAA29079..1; -.
CC KW Antigen; Sporozoite;. Repeat; Sporulation.
CC FT NON_TER 1
CC FT DOMAIN 18 210 12 X APPROXIMATE TANDEM REPEATS, GLN-RICH.
CC FT REPEAT 18 48 1.
CC FT REPEAT 49 57 2.
CC FT REPEAT 58 65 3.
CC FT REPEAT 66 78 4.
CC FT REPEAT 79 90 5.
CC FT REPEAT 91 103 6.
CC FT REPEAT 104 140 7.
CC FT REPEAT 141 152 8.
CC FT REPEAT 153 164 9.
CC FT REPEAT 165 172 10.
```

```

FT REPEAT 173 192 11.
FT REPEAT 193 210 12.
FT NON_TER 255 255
SQ SEQUENCE 255 AA; 31267 MW; 8C5E605FFFC2DB3 CRC64;

Query Match 25.1%; Score 86; DB 1; Length 255;
Best Local Similarity 32.3%; Pred. No. 0.075;
Matches 20; Conservative 16; Mismatches 22; Indels 4; Gaps 2;

OY 2 OROVOC--OGRCOPGOGGOREOOCORCKMEQYKGEJGEH--ENYNHKKNSREEEG 57
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 123 OQQLQCCGQ0000000LQ0QWSE00000000QMPPE0000000WFE00000QMSD00000A 182

OY 58 Q0 59
   ||
Db 183 Q0 184

RESULT 10
Y192_HUMAN STANDARD; PRT; 2124 AA.
AC 093074;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA0192 (FRAGMENT).
GN KIAA0192.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
   (1)
RN RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW.
RX MEDLINE=96281124; PubMed=8724849;
RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (K1A0161-K1A0200) deduced by
RL analysis of cDNA clones from human cell line KG-1."
DNA Res. 3:17-24(1996).
-1- TISSUE SPECIFICITY: UBICUITOUS.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D83783; BAA12112.1; -.
DT Hypothetical protein.
FT NON_TER 1
FT DOMAIN 599 602 POLY-SER.
FT DOMAIN 1201 1207 POLY-GLY.
FT DOMAIN 1998 2124 GLN-RICH.
FT DOMAIN 1998 2023 POLY-GLN.
FT DOMAIN 2028 2033 POLY-GLN.
FT DOMAIN 2037 2070 POLY-GLN.
FT DOMAIN 2090 2097 POLY-GLN.
SQ SEQUENCE 2124 AA; 237207 MW; 255FB9419EC39F42 CRC64;

Query Match 25.1%; Score 86; DB 1; Length 2124;
Best Local Similarity 30.5%; Pred. No. 0.49;
Matches 18; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

OY 2 OROVOCGRCGROEOGOGGOREOOCORCKMEQYKGEJGEH--ENYNHKKNSREEEGOR 60
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 1999 Q000000000000000000000000YHNR000000ILR000000000000000000000 2057

RESULT 11

```

ID	PRO_DROME	STANDARD;	PRT;	1403 AA.
AC	P29617;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	PROTEIN PROSPERO.			
GN	PROSPERO.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephyridioidea; Neophyllidae; Drosophila.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92069760; Pubmed=1720353;			
RA	Vassalin H., Grell E., Wolff E., Blier E., Jan L.Y., Jan Y.N.;			
RT	"Prospero is expressed in neuronal precursors and encodes a nuclear			
RT	protein that is involved in the control of axonal outgrowth in			
RT	drosophila."			
RL	Cell 67:941-953(1991).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92171948; Pubmed=1540176;			
RA	Matsuzaki F., Kotzumi K., Hama C., Yoshitaka T., Nabeshima Y.;			
RT	"Cloning of the Drosophila prospero gene and its expression in			
RT	ganglion mother cells."			
RL	Biochem. Biophys. Res. Commun. 182:1326-1332(1992).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93083413; Pubmed=1842358;			
RA	Chu-Lagraff O., Wright D.M., McNeill L.K., Doe C.O.;			
RT	"The prospero gene encodes a divergent homeodomain protein that			
RT	controls neuronal identity in Drosophila."			
RL	Development Suppl. 2:79-85(1991).			
RN	[4]			
RP	SIMILARITY TO C.ELEGANS CERN-26.			
RX	MEDLINE=94212446; Pubmed=7909177;			
RA	Buerklin T.R.;			
RL	"A Caenorhabditis elegans prospero homologue defines a novel domain."			
RL	Trends Biochem. Sci. 19:70-71(1994).			
CC	-I- FUNCTION: INVOLVED IN THE CONTROL OF OTHER NEURONAL PRECURSOR			
CC	GENES AS WELL AS AXONAL OUTGROWTH AND PATHFINDING OF NUMEROUS			
CC	CENTRAL AND PERIPHERAL NEURONS. IT IS PROBABLY GENERALLY REQUIRED			
CC	FOR PROPER NEURONAL DIFFERENTIATION OF MOST OR ALL NEURONS & THEIR			
CC	PRECURSORS IN THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS, BUT NOT			
CC	FOR THE SPECIFICATION OF IDENTITY. PROSPERO PROTEIN MAY REGULATE			
CC	TRANSCRIPTION BY BINDING TO DNA.			
CC	-I- SUBCELLULAR LOCATION: NUCLEAR.			
CC	-I- SUBCELLULAR LOCATION: NUCLEAR.			
CC	-I- ALTERNATIVE PRODUCTS: 2 ISOFORMS; PROS-L (SHOWN HERE) AND PROS-			
CC	S; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY HAVE DIFFERENT N-			
CC	TERMINAL AMINO ACIDS OF THE HOMODOMAIN.			
CC	-I- TISSUE SPECIFICITY: NEURONAL PRECURSORS. EXPRESSED IN THE			
CC	DEVELOPING CNS, LENS-SECRETING CONE CELLS OF THE EYE, AND MIDGUT.			
CC	-I- DEVELOPMENTAL STAGE: EXPRESSED IN NEURONAL PRECURSORS EARLY DURING			
CC	FORMATION.			
CC	-I- SIMILARITY: BELONGS TO THE PROSPERO FAMILY OF HOMEOBOX PROTEINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M81389; AAA28841.1; -			
DR	EMBL; D10609; BAA01464.1; -			
DR	EMBL; Z11743; CAAT7802.1; -			
DR	PIR; A41089; A41089.			
DR	PIR; J01397; J01397.			
DR	FLYBASE; FBgn0004595; pros			
DR	Nuclear protein; Transcription regulation; DNA-binding; Homeobox;			
DR	Developmental protein; Alternative splicing.			
DR	-----			

FT	DOMAIN	4	12	POLY-ALA.				
FT	DOMAIN	28	31	POLY-SER.				
FT	DOMAIN	32	35	POLY-ASN.				
FT	DOMAIN	188	191	POLY-ALA.				
FT	DOMAIN	253	260	POLY-GLN.				
FT	DOMAIN	270	276	POLY-GLN.				
FT	DOMAIN	282	286	POLY-ASN.				
FT	DOMAIN	431	437	POLY-ASP.				
FT	DOMAIN	505	508	POLY-ALA.				
FT	DOMAIN	717	737	POLY-GLN.				
FT	DOMAIN	754	763	POLY-GLN.				
FT	DOMAIN	766	772	POLY-GLN.				
FT	DOMAIN	934	937	POLY-ALA.				
FT	DOMAIN	952	957	POLY-ALA.				
FT	DOMAIN	960	963	POLY-GLN.				
FT	DOMAIN	966	970	POLY-GLN.				
FT	DOMAIN	991	998	NUCLEAR LOCALIZATION SIGNAL.				
FT	DOMAIN	1029	1048	POLY-GLN.				
FT	DOMAIN	1074	1082	POLY-PRO.				
FT	DNA_BIND	1241	1303	HOMEOBOX (ATYPICAL).				
FT	DOMAIN	1304	1403	PROSPERO-LIKE.				
FT	VARSPLIC	1216	1244	MISSED (IN ISOFORM PROS-S).				
FT	CONFLICT	76	98	AKMINTEFORKODMTSGIP -> GODERAIVMPPEDE				
FT	CONFLICT			GGRNRWPA (IN REF. 1).				
FT	CONFLICT	120	144	ISLWTSKLIOQNNNSIAPAS -> NLALOFHYOVAR				
FT	CONFLICT			AAAIITALLPPIG (IN REF. 1).				
FT	CONFLICT	418	418	H -> Q (IN REF. 2).				
FT	CONFLICT	677	677	A -> C (IN REF. 1).				
FT	CONFLICT	802	802	A -> S (IN REF. 1).				
FT	CONFLICT	958	958	T -> S (IN REF. 1).				
SO	SEQUENCE	1403 AA;	153569 MW;	Q -> QQQQQ (IN REF. 1).				
				9EFB973E2A2E38E CRC64;				
Query Match		24.9%;	Score 85;	DB 1;				
Best Local Similarity		32.7%;	Pred. No. 0.42;	Length 1403;				
Matches 18;		Conservative 17;	Mismatches 20;	Indels 0;				
				Gaps				
Oy	6 OQCQGRCEQDQGREDQCCQRCKMGEYKEOEFGHEHNTNHKKRSEEEGCGOR 60							
	: : : : : : : : : : :							
Db	717 QQQQQQQQQQQQQQQQQQQQQEQQRREDEQEQQRKRKEEDQQQIQRDQQLDLQQQ 771							
RESULT 12								
KAPC_DICDI								
ID KAPC_DICDI	STANDARD:	PRT:	648 AA.					
AC P34099;								
DT 01-FEB-1994	(Rel. 28,	Created)						
RA Buertl E., Anjard C., Schodler J.-C., Raymond C.D.;								
RT "Isolation of two genes encoding putative protein kinases regulated								
DE CAMP-DEPENDENT PROTEIN KINASE CATALYTIC SUBUNIT (EC 2.7.1.37).								
CN PRAC OR PK2 OR PK3.								
OS Dictyostelium discoideum (Slime mold).								
OC Eukaryota; Dictyostelida; Slime mold.								
RN [1]								
RP SEQUENCE FROM N.A.								
RX MEDLINE=91323730;	PubMed=1864510;							
RA Buertl E., Anjard C., Schodler J.-C., Raymond C.D.;								
RT "Isolation of two genes encoding putative protein kinases regulated								
during Dictyostelium discoideum development."								
RL Gene 102:57-65(1991).								
RN [2]								
RP CHARACTERIZATION.								
RX MEDLINE=93385090;	PubMed=8373760;							
RA Anjard C., Etcheberry L., Pinard S., Veron M., Raymond C.D.;								
RT "An unusual catalytic subunit for the CAMP-dependent protein kinase								
of Dictyostelium discoideum."								
RL Biochemistry 32:9532-9538(1993).								
RN [3]								
RP CHARACTERIZATION.								
RC STRAIN=AX3;								
RX MEDLINE=93066311;	PubMed=1332055;							
RA Mann S.K.O., Yonemoto W.M., Taylor S.S., Firtel R.A.;								

```

RT "ddPK3, which plays essential roles during Dictyostelium development,
RT encodes the catalytic subunit of cAMP-dependent protein kinase."
RL Proc. Natl. Acad. Sci. U.S.A. 89:10701-10705(1992).
CC -1- FUNCTION: ESSENTIAL FOR DIFFERENTIATION AND FRUIT MORPHOGENESIS.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN = ADP + A PHOSPHOPROTEIN.
CC -1- SUBUNIT: IN DICTYOSTELIUM THE HOLOENZYME IS A DIMER COMPOSED OF
CC A REGULATORY (R) AND A CATALYTIC (C) SUBUNIT. IN THE PRESENCE OF
CC CAMP IT DISSOCIATES INTO THE ACTIVE C SUBUNIT AND AN R MONOMER.
CC -1- DEVELOPMENTAL STAGE: CAMP ACTIVITY IS LOW IN VEGETATIVELY GROWING
CC AMOEBAE, INCREASES DURING DEVELOPMENT OF AGGREGATION AND REACHES
CC A MAXIMUM AT CULMINATION.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CAMP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M38703; -; NOT_ANNOTATED_CDS.
DR PIR: J01150.
DR HSSP: P05132; ZCPK.
DR DICTYDB: DD02030; PKAC.
DR INTERPRO: IPR000719; -.
DR INTERPRO: IPR000961; -.
DR INTERPRO: IPR002290; -.
DR PRAM: PF00069; pkinase_1.
DR PRAM: PF00433; pkinase_C_1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP_1.
DR PROSITE: PS00108; PROTEIN_KINASE_SF_1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM_1.
DR Transferrase: Serine/threonine-protein kinase; ATP-binding; CAMP;
DR Phosphorylation.
KM DOMAIN 58 64 ASN-RICH.
FT DOMAIN 136 223 GLN-RICH.
FT DOMAIN 233 250 THR-RICH.
FT DOMAIN 336 590 PROTEIN KINASE.
FT NP_BIND 342 350 ATP (BY SIMILARITY).
FT BINDING 365 365 ATP (BY SIMILARITY).
FT ACT_SITE 459 459 BY SIMILARITY.
FT MOD_RES 490 490 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 648 AA; 74458 MW; D0F9B3448C58D084 CRC64;

Query Match 24.7%; Score 84.5; DB 1; Length 648;
Best Local Similarity 40.4%; Pred. No. 0.24;
Matches 19; Conservative 10; Mismatches 13; Indels 5; Gaps 1;

QY 2 GROYQCCGRCGCEQOOCGRQOOCORCKCWEQYGEQGEHNNHKK 48
DB 186 QLOQQQLQOQLQOQQQQQQQQQQQQ-----QOKOQKQOQQQQOHHQ 227

RESULT 13
RBLF CANAL
ID RBLF CANAL STANDARD; PRT; 527 AA.
AC Q00312;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TRANSCRIPTION FACTOR RBLF (RPG-BOX-BINDING FACTOR) (REPRESSOR-
DE ACTIVATOR PROTEIN 1).
GN RBLF
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC anamorphic Saccharomycetales; Candida.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 10231;
RX MEDLINE=97195789; PubMed=9043119;

```

```

RA Ishii N., Yamamoto M., Lahm H.-W., Iizumi S., Yoshinara F.,
RA Nakayama H., Aisawa M., Aoki Y.;
RT "A DNA-binding protein from Candida albicans that binds to the RPG
RT box of Saccharomyces cerevisiae and the telomeric repeat sequence of
RT C. albicans."
RL Microbiology 143:417-427(1997).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=97195789; PubMed=9043120;
RA Ishii N., Yamamoto M., Yoshinara F., Aisawa M., Aoki Y.;
RT "Biochemical and genetic characterization of Rbf1p, a putative
RT transcription factor of Candida albicans."
RL Microbiology 143:429-435(1997).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS TO THE RPG BOX AND
CC TO TELOMERES. MAY BE INVOLVED IN THE REGULATION OF THE TRANSITION
CC BETWEEN YEAST AND FILAMENTOUS FORMS AT THE LEVEL OF TRANSCRIPTION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; PREDOMINANTLY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D85862; BAAL2888.1; -.
DR Transcription regulation; DNA-binding; Activator; Nuclear protein;
KM Telomere.
FT DNA_BIND 160 300 POTENTIAL.
FT DOMAIN 73 80 POLY-ALA.
FT DOMAIN 135 139 POLY-GLN.
FT DOMAIN 326 329 POLY-HIS.
FT DOMAIN 332 353 POLY-GLN.
FT DOMAIN 398 405 POLY-GLN.
FT DOMAIN 445 452 POLY-ALA.
FT DOMAIN 478 481 POLY-GLN.
SQ SEQUENCE 527 AA; 59441 MW; A0DD675F937A807D CRC64;

Query Match 24.6%; Score 84; DB 1; Length 527;
Best Local Similarity 30.6%; Pred. No. 0.22;
Matches 19; Conservative 11; Mismatches 24; Indels 8; Gaps 1;

QY 3 ROYQCCGRCGCEQOOCGRQOOCORCKCWEQYGEQGEHNNHKK 54
DB 108 QOYRQAGAQAGAQAGQAGRHQQLQAHAYQQOQLHQLGQLSQQLAQPHLSQEHYRDLTT 167

QY 55 EE 56
DB 168 DE 169

RESULT 14
INVO_AOTTR
ID INVO_AOTTR STANDARD; PRT; 544 AA.
AC P24708;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE INVOLUCRIN.
GN IVL.
OS Aotus trivirgatus (Night monkey) (Douroucouli).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90014142; PubMed=2507864;
RA Tseng H., Green H.;
RT "The involucrin gene of the owl monkey: origin of the early region."
RL Mol. Biol. Evol. 6:460-468(1989).
CC -1- FUNCTION: INVOLUCRIN IS A KERATINOCTE PROTEIN THAT FIRST APPEARS
CC IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO

```


RP [14]
 RP VARIANT LNCAP ALA-877.
 RX MEDLINE=91083633; PubMed=2260966.
 RA Veldscholte J., Rüs-Stalpers C., Kuiper G.G., Jenster G.,
 RA Berrevoets C., Claassen E., Van Rooij H.C.J., Trapman J.,
 RA Brinkman A.O., Mulder E.;
 RT "A mutation in the ligand binding domain of the androgen receptor of
 RT human LNCAP cells affects steroid binding characteristics and
 RT response to anti-androgens";
 RL Biochem. Biophys. Res. Commun. 173:534-540(1990).
 RN [15]
 RP VARIANT CAIS MET-866.
 RX MEDLINE=91186983; PubMed=2082179;
 RA Brown T.R., Lubahn D.B., Wilson E.M., French F.S., Migson C.J.,
 RA Corfen J.L.;
 RT "Functional characterization of naturally occurring mutant androgen
 RT receptors from subjects with complete androgen insensitivity.";
 RL Mol. Endocrinol. 4:1759-1772(1990).
 RN [16]
 RP VARIANT CYS-774.
 RX MEDLINE=91310758; PubMed=1856263;
 RA MacCalli M., Tilley W.D., Zoppi S., Griffin J.E., Wilson J.D.,
 RA McPhall M.J.;
 RT "Androgen resistance associated with a mutation of the androgen
 RT receptor at amino acid 772 (Arg-->Cys) results from a combination of
 RT decreased messenger ribonucleic acid levels and impairment of
 RT receptor function";
 RL J. Clin. Endocrinol. Metab. 73:318-325(1991).
 RN [17]
 RP VARIANT CAIS ASN-695 AND HIS-695, AND SEQUENCE OF 629-723 FROM N.A.
 RX MEDLINE=92131007; PubMed=1775137;
 RA Rüs-Stalpers C., Trifiro M.A., Kuiper G.G., Jenster G., Romalo G.,
 RA Sai T., Van Rooij H.C., Kaufman M., Rosenfield R.L., Liao S.;
 RT "Substitution of aspartic acid-686 by histidine or asparagine in the
 RT human androgen receptor leads to a functionally inactive protein with
 RT altered hormone-binding characteristics.";
 RL Mol. Endocrinol. 5:1562-1569(1991).
 RN [18]
 RP VARIANTS CAIS AND PAIS.
 RX MEDLINE=93338440; PubMed=1307250;
 RA Hatch J.A., Williams D.M., Davies H.R., Brown B.D., Evans B.A.J.,
 RA Hughes J.A., Patterson M.N.;
 RT "Androgen receptor gene mutations identified by SSCP in fourteen
 RT subjects with androgen insensitivity syndrome";
 RL Hum. Mol. Genet. 1:497-503(1992).
 RN [19]
 RP VARIANT CAIS VAL-787.
 RX MEDLINE=92235226; PubMed=1569163;
 RA Nakao R., Hagi M., Yanase T., Ogo A., Takayanagi R., Katsube T.,
 RA Fukunaki Y., Nawata H.;
 RT "A single amino acid substitution (Met-786-->Val) in the steroid-
 RT binding domain of human androgen receptor leads to complete androgen
 RT insensitivity syndrome.";
 RL J. Clin. Endocrinol. Metab. 74:1152-1157(1992).
 RN [20]
 RP VARIANT LNCAP ALA-877.
 RX MEDLINE=92222955; PubMed=1562539;
 RA Veldscholte J., Berrevoets C.A., Rüs-Stalpers C., Kuiper G.G.,
 RA Jenster G., Trapman J., Brinkman A.O., Mulder E.;
 RT "The androgen receptor in LNCAP cells contains a mutation in the
 RT ligand binding domain which affects steroid binding characteristics
 RT and response to antiandrogens";
 RL J. Steroid Biochem. Mol. Biol. 41:665-669(1992).
 RN [21]
 RP VARIANT MET-730.
 RX MEDLINE=92335289; PubMed=1631125;
 RA Newmark J.R., Hardy D.O., Tonb D.C., Carter B.S., Epstein J.L.,
 RA Isaacs W.B., Brown T.R., Trapack E.R.;
 RT "Androgen receptor gene mutations in human prostate cancer.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:6319-6323(1992).
 RN [22]
 RP VARIANT CAIS VAL-754.
 RX MEDLINE=93372806; PubMed=8103396;

RA Llobrocacia J.-M., Lumbroso S., Ktari R., Dunas R., Sultan C. ;
RT "A exonic point mutation creates a Maellii site in the androgen
RT receptor gene of a family with complete androgen insensitivity
RT syndrome." ;
RL Hum. Mol. Genet. 2:1041-1043(1993) .
RN [23]
RP VARIANT CAIS ARG-807 .
RX MEDLINE=94108430; PubMed=8281140;
RA Adevemo O., Kallio P.J., Palvimo J.J., Kontula K., Jaenke O.A. ;
RT "A single-base substitution in exon 6 of the androgen receptor gene
RT causing complete androgen insensitivity: the mutated receptor fails
RT to transactivate but binds to DNA in vitro." ;
RL Hum. Mol. Genet. 2:1809-1812(1993) .
RN [24]
RP VARIANT PAIS VAL-743 .
RX MEDLINE=93315568; PubMed=8325932;
RA Nakao R., Yanase T., Sakai Y., Hagi M., Nawata H. ;
RT "A single amino acid substitution (Gly743 --> Val) in the steroid-
RT binding domain of the human androgen receptor leads to Reifenshtein

```

Query Match      24.4%; Score 83.5; DB 1; length 919;
Best Local Similarity 32.3%; Pred. No. 0.4;
Matches 21; Conservative 17; Mismatches 16; Indels 11; Gaps 2;

QY      1 LQRYQCCGCGRCQEQEQGCRQEQCCQCRKCEQDYKQERGEHNNHKKNS-----EE 54
      ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||
Db      57 LQQQQQQQ-----QQQQQQQQQQQQQQQQQQQETSPRQQQQQQGEGDSQQAHRKPGTGLVLEE 111
      ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||

QY      55 EEGQQ 59
      :::|
Db     112 QQPSQ 116

```

```
Search completed: March 1, 2001, 16:16:54
Job time: 434 sec
```

